Applied Probability Trust (8 December 2009)

STEIN'S METHOD AND STOCHASTIC ORDERINGS

FRASER DALY,[*] *Universität Zürich*

CLAUDE LEFÈVRE,[**] *Université Libre de Bruxelles*

SERGEY UTEV,[***] *University of Nottingham*

Abstract

A stochastic ordering approach is applied with Stein's method for approximation by the equilibrium distribution of a birth–death process. The usual stochastic order and the more general s-convex orders are discussed. Attention is focused on Poisson and translated Poisson approximation of a sum of dependent Bernoulli random variables, for example k–runs in i.i.d. Bernoulli trials. Other applications include approximation by polynomial birth–death distributions.

Keywords: Stein's method; birth–death process; stochastic ordering; total variation distance; (in)dependent indicators; (translated) Poisson approximation; total negative and positive dependence; (approximate) local dependence; polynomial birth–death approximation; k–runs

2000 Mathematics Subject Classification: Primary 62E17

Secondary 60F05; 60J80

1. Introduction

Stein's method has proved to be an effective tool in probability approximation, and has the advantage of being applicable in the presence of dependence. See, for

[*] Postal address: Institut für Mathematik, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. Email address: fraser.daly@math.uzh.ch

[**] Postal address: Université Libre de Bruxelles (U.L.B.), Département de Mathématique, Campus de la Plaine C.P. 210, B-1050 Bruxelles, Belgium. Email address: clefevre@ulb.ac.be

[***] Postal address: University of Nottingham, School of Mathematical Sciences, University Park, NG7 2RD Nottingham, United Kingdom. Email address: sergey.utev@nottingham.ac.uk

example, Stein (1986), and Barbour and Chen (2005) for more recent developments. It is well–known that error bounds obtained via Stein's method may be simplified under some assumptions on the dependence present. For example, in the presence of negative or positive relation, Stein's method gives simple error bounds in the Poisson approximation of a sum of indicator random variables. This is exploited throughout the work of Barbour *et al.* (1992), and will be returned to in our Section 4.

In this work, we consider the more general situation of approximation by the equilibrium distribution of a birth–death process, and examine the situations in which Stein's method leads to simple, easily calculable error bounds. These error bounds will typically be differences of moments of our random variables. As we will see, the assumptions under which we can obtain such error bounds are naturally phrased in terms of stochastic orderings.

Consider a birth-death process on (some subset of) $\mathbb{Z}^+$ with birth rates α_j and death rates β_j for $j \geq 0$. Suppose $\beta_0 = 0$. Let π be the stationary distribution of such a process, with $\pi_j = P(\pi = j)$, $j \geq 0$. In this work we combine Stein's method with a stochastic ordering construction to consider the approximation by π of some random variable W on $\mathbb{Z}^+$.

Our random variable π satisfies the identity $E[Ag(\pi)] = 0$ for any bounded function $g : \mathbb{Z}^+ \mapsto \mathbb{R}$, where A is the linear operator defined by

$$Ag(j) = \alpha_j g(j+1) - \beta_j g(j), \quad j \geq 0. \tag{1}$$

A is a characterising operator for π, in the sense that a random variable $Z =_d \pi$ if and only if $E[Ag(Z)] = 0$ for all g bounded. The construction of such a characterising operator is the basis of Stein's method for probability approximation. See the books by Stein (1986), Barbour *et al.* (1992), Barbour and Chen (2005) and references therein. For Stein's method applied to birth-death processes, see Brown and Xia (2001) and Holmes (2004).

Given some test function h, the so-called Stein equation is defined by

$$h(j) - E[h(\pi)] = Af(j), \quad j \geq 0. \tag{2}$$

Its solution is denoted $f = f_h = Sh$. We call S the Stein operator. Bounds on S are an essential ingredient of Stein's method.

Note that the solution f of the Stein equation depends on the chosen test function h. However, for notational convenience in much of the work that follows we will write f rather than f_h or Sh. We will often choose $h(j) = I_{(j\in B)}$ for some $B \subseteq \mathbb{Z}^+$, in which case the solution f will depend on the chosen set B.

There are several common distributions π covered by this framework. For each of the examples below, bounds are available on the corresponding Stein operator S. Theorem 2.10 of Brown and Xia (2001) may also be applied to give bounds on S in many cases.

- If $\alpha_j = \lambda$ and $\beta_j = j$, then $\pi \sim \mathrm{Po}(\lambda)$, the Poisson distribution with mean λ. See Barbour *et al.* (1992) and references therein.
- If $\alpha_j = q(r+j)$ and $\beta_j = j$, then $\pi \sim \mathrm{NB}(r, 1-q)$ has a negative binomial distribution. See Brown and Phillips (1999).
- If $\alpha_j = (n-j)p$ and $\beta_j = (1-p)j$, then $\pi \sim \mathrm{Bin}(n,p)$. See Ehm (1991).
- In the geometric case, we may, of course, use the negative binomial operator above. Alternatively we may choose $\alpha_j = q$ and $\beta_j = I_{(j\geq 1)}$, so that $\pi \sim \mathrm{Geom}(1-q)$. See Peköz (1996).

The present work is organized as follows. In Section 2, we will derive abstract error bounds using Stein's method combined with some stochastic ordering assumptions in the setting of approximation by the equilibrium distribution of a birth–death process. In Section 3, a simple sufficient condition under which these stochastic ordering assumptions hold is considered, and some applications are given. Section 4 discusses Poisson approximation for a sum of dependent indicators. We will see how concepts of negative and positive relation relate to our stochastic ordering assumptions, and present generalizations of error bounds derived by Barbour *et al.* (1992). Based on this work we move on, in Section 5, to consider translated Poisson approximation. Applications here will include approximation of the number of k–runs in i.i.d. Bernoulli trials. Finally, in Section 6, we give another abstract approximation theorem, and consider its application to a sum of independent indicator random variables.

2. An abstract approximation theorem

Consider Stein's method for approximating the equilibrium distribution of a birth-death process. Our purpose in this section is to derive abstract error bounds under some stochastic ordering assumptions.

2.1. A first-order bound

Suppose that W is a random variable supported on (some subset of) $\mathbb{Z}^+$ with $\mu_j = P(W = j)$, $j \geq 0$. Set $\mu_{-1} = 0$. Our concern is the approximation of such a variable W by π, specifically by estimating the difference $|Eh(W) - Eh(\pi)|$, i.e. $|E[\mathcal{A}f(W)]|$. For this, a simple representation of this difference will be applied with some stochastic ordering assumptions to yield bounds using Stein's method. We may then bound, for example, the total variation distance between $\mathcal{L}(W)$ and $\mathcal{L}(\pi)$, defined by

$$d_{TV}(\mathcal{L}(W), \mathcal{L}(\pi)) = \sup_{B \subseteq \mathbb{Z}^+} |P(W \in B) - P(\pi \in B)|.$$

Although we are mainly concerned with approximation in total variation distance, the results we derive may also be used with other probability metrics.

Let Δ be the forward difference operator. Since, with the operator (1), the choice of $f(0)$ is arbitrary, we follow Brown and Xia (2001) and choose $f(0) = 0$. Writing $f(j) = \Delta f(0) + \cdots + \Delta f(j-1)$, we thus obtain the representation

$$Eh(W) - Eh(\pi) = \sum_{k=0}^{\infty} \Delta f(k) \sum_{j=k+1}^{\infty} (\alpha_{j-1}\mu_{j-1} - \beta_j \mu_j). \tag{3}$$

In the next subsection, we will extend (3) to include the lth forward differences of $f(\cdot)$, for all $l \geq 1$.

We now consider how this representation may be applied in conjunction with the usual stochastic ordering, denoted $\succeq_{st}$. Define two random variables W_α and W_β by

$$P(W_\alpha = j) = \frac{\alpha_{j-1}\mu_{j-1}}{E\alpha_W}, \quad \text{and} \quad P(W_\beta = j) = \frac{\beta_j \mu_j}{E\beta_W}, \quad j \geq 1. \tag{4}$$

If $W_\alpha \succeq_{st} W_\beta$ and $E\alpha_W \geq E\beta_W$, we have that $\sum_{j=i}^{\infty} \alpha_{j-1}\mu_{j-1} \geq \sum_{j=i}^{\infty} \beta_j \mu_j$ for all $i \geq 1$. In this case, (3) may be bounded to obtain

$$|Eh(W) - Eh(\pi)| \;\leq\; \|\Delta f\|_\infty \, E[\alpha_W(W+1) - \beta_W W].$$

A similar argument holds if we instead assume that $W_\beta \succeq_{st} W_\alpha$ and $E\beta_W \geq E\alpha_W$. We thus obtain the following result.

Proposition 1. *Assume that one of the two following conditions holds:*

$$\textit{either (i) } W_\alpha \succeq_{st} W_\beta \ \textit{ with } E\alpha_W \geq E\beta_W, \ \textit{ or (ii) } W_\beta \succeq_{st} W_\alpha \ \textit{ with } \ E\beta_W \geq E\alpha_W. \tag{5}$$

Then,

$$|Eh(W) - Eh(\pi)| \ \leq \ \|\Delta Sh\|_\infty \ |E[\alpha_W(W+1) - \beta_W W]| \, . \tag{6}$$

2.2. A s-order bound

We will now establish our main abstract result. For that, we will have recourse to the concept of discrete s-convex stochastic ordering, denoted $\succeq_{s-cx}$, for any integer $s \geq 1$. See, for example, Lefèvre and Utev (1996) for this notion. Briefly, given any two non-negative integer-valued random variables X and Y, one says that $X \succeq_{s-cx} Y$ when

$$E[f(X)] \leq E[f(Y)] \ \text{ for all } s\text{-convex functions } f,$$

that is, for all functions f satisfying $\Delta^s f(j) \geq 0, \ j \geq 0$. Note that this ordering implies that X and Y have the same first $s-1$ moments.

To begin with, we introduce a Bernoulli random variable v_p with

$$P(v_p = 1) = p = 1 - P(v_p = 0),$$

independently of all other entries. We write $\alpha = E\alpha_W$, $\beta = E\beta_W$, and in an analogous way to (4), we define the random variables W_α and W_β by

$$P(W_\alpha \in B) = \alpha^{-1} E[\alpha_W I_{(W+1 \in B)}], \ \text{ and } \ P(W_\beta \in B) = \beta^{-1} E[\beta_W I_{(W \in B)}], \tag{7}$$

for any Borel set B. For notational convenience, we choose to write $C_n^k = \binom{n}{k}$.

The key theorem and an immediate corollary will be first stated, the proof of the theorem being given after.

Proposition 2. *Assume that there exists a random variable Y on $\mathbb{Z}^+$ such that $W_\beta - Y \geq 0$ a.s. and*

$$W_\alpha \succeq_{s-cx} v_p(W_\beta - Y). \tag{8}$$

Then,

$$|Eh(W) - Eh(\pi)| \leq \sum_{t=0}^{s-1} |\Delta^t Sh(0)| \, |E(\alpha_W C^t_{W+1}) - E(\beta_W C^t_W)|$$
$$+ \|\Delta^s Sh\|_\infty \, (\alpha E[C^s_{W_\alpha}] - 2\alpha p E[C^s_{W_\beta - Y}] + (\alpha p + |\alpha p - \beta|) E[C^s_{W_\beta}]). \quad (9)$$

Consider the special case of (8) when $p = 1$ and $Y = 0$ a.s. When $\alpha = \beta$ and under the condition (10) below, one has that

$$E[\alpha_W (W+1)^t] = E[\beta_W W^t], \quad t = 0, \ldots, s-1,$$

so that the inequality (9) reduces to (11).

Corollary 1. *Assume that $\alpha = \beta$, and one of the two following conditions holds:*

$$\textit{either (i)} \;\; W_\alpha \succeq_{s-cx} W_\beta, \quad \textit{or (ii)} \;\; W_\beta \succeq_{s-cx} W_\alpha. \quad (10)$$

Then,

$$|Eh(W) - Eh(\pi)| \;\leq\; \|\Delta^s Sh\|_\infty \, |E[\alpha_W C^s_{W+1}] - E[\beta_W C^s_W]|. \quad (11)$$

We note that Proposition 1 does not follow as a special case of Corollary 1, since this latter result requires the condition $\alpha = \beta$ not needed in Proposition 1.

Proof of Proposition 2. In the first step we derive a representation of $E[Af(W)]$ that generalizes the representation (3). Observe that (1) and (7) give

$$E[Af(W)] = E[\alpha_W f(W+1)] - E[\beta_W f(W)] = \alpha E[f(W_\alpha)] - \beta E[f(W_\beta)].$$

Expanding the function f by the discrete Taylor formula, we obtain, for any $s = 1, 2, \ldots,$

$$f(x) = f(0) + \sum_{k=0}^{\infty} \Delta f(k) \, I_{(x>k)} = \sum_{t=0}^{s-1} \Delta^t f(0) \, C^t_x \; + \; \sum_{k=0}^{\infty} \Delta^s f(k) \, C^{s-1}_{x-k-1};$$

see Lefèvre and Utev (1996). Thus, we find that

$$\begin{aligned} E[Af(W)] &= \sum_{t=0}^{s-1} \Delta^t f(0) \, E[AC^t_W] + \sum_{k=0}^{\infty} \Delta^s f(k) \, E[AC^{s-1}_{W-k-1}] \\ &= \sum_{t=0}^{s-1} \Delta^t f(0) \, (\alpha E[C^t_{W_\alpha}] - \beta E[C^t_{W_\beta}]) \\ &\quad + \sum_{k=0}^{\infty} \Delta^s f(k) \, (\alpha E[C^{s-1}_{W_\alpha - k-1}] - \beta E[C^{s-1}_{W_\beta - k-1}]). \end{aligned} \quad (12)$$

Our next step is to derive an abstract metrics-ordering relationship result, which is stated below as a separate lemma. Using the bound (14) in the representation (12) then leads to the announced bound (11).

Lemma 1. *Let X, Y and Z be random variables on $\mathbb{Z}^+$ such that*

$$Z - Y \geq 0 \quad a.s., \quad \text{and} \;\; X \succeq_{s-cx} v_p(Z - Y). \tag{13}$$

Then, for all $a, b \in \mathbb{R}^+$,

$$\sum_{k=0}^{\infty} |aE[C^{s-1}_{X-k-1}] - bE[C^{s-1}_{Z-k-1}]| \;\leq\; aE[C^s_X] - 2apE[C^s_{Z-Y}] + (ap + |ap - b|)E[C^s_Z]. \tag{14}$$

Proof. Letting

$$w_k^{(s)}(x) = w_k(x) = C^{s-1}_{x-k-1},$$

we get that

$$\begin{aligned}
\sum_{k=0}^{\infty} |aE(C^{s-1}_{X-k-1}) - bE(C^{s-1}_{Z-k-1})| \;&=\; \sum_{k=0}^{\infty} |aE[w_k(X)] - bE[w_k(Z)]| \\
\leq a \sum_{k=0}^{\infty} |E[w_k(X)] - E[w_k(v_p(Z-Y))]| &+ a \sum_{k=0}^{\infty} |E[w_k(v_p Z)] - E[w_k(v_p(Z-Y))]| \\
+ \sum_{k=0}^{\infty} |aE[w_k(v_p Z)] - bE[w_k(Z)]| \;&=\; S_1 + S_2 + S_3.
\end{aligned} \tag{15}$$

Let us examine the three sums in (15). First, we easily check that

$$\sum_{k=0}^{\infty} E[w_k(Z)] = E[C^s_Z]. \tag{16}$$

Using (16), we successively find that

$$S_3 \;=\; |ap - b| \sum_{k=0}^{\infty} E[w_k(Z)] \;=\; |ap - b| E[C^s_Z];$$

since $Z - Y \geq 0$ and $Z \succeq_{st} Z - Y$,

$$S_2 \;=\; ap \sum_{k=0}^{\infty} (E[w_k(Z)] - E[w_k(Z - Y)]) \;=\; ap(E[C^s_Z] - E[C^s_{Z-Y}]);$$

finally, by the assumption (13) and a standard property of the order $\succeq_{s-cx}$,

$$S_1 \;=\; a \sum_{k=0}^{\infty} [Ew_k(X) - pEw_k(Z - Y)] \;=\; a(E[C^s_X] - pE[C^s_{Z-Y}]).$$

Inserting these three terms in (15), we then deduce the bound (14).

Remark 1. For $s = p = 1$ and $a = b = 1$, Lemma 1 states that if $X \succeq_{st} Z - Y \geq 0$, then an upper bound for the Wasserstein distance between $\mathcal{L}(X)$ and $\mathcal{L}(Z)$ is

$$d_W(\mathcal{L}(X), \mathcal{L}(Z)) \;=\; \sum_{k=0}^{\infty} |P(X > k) - P(Z > k)| \;\leq\; 2EY + EX - EZ. \tag{17}$$

This bound is of interest in the stochastic ordering context investigated by Kamae *et al.* (1977), with random variables on $\mathbb{Z}^+$ here. Note that by choosing the optimal coupling X, Z and $Y = (Z - X)_+$, (17) gives the exact bound since

$$d_W(\mathcal{L}(X), \mathcal{L}(Z)) \;\leq\; 2E(Z - X)_+ + EX - EZ \;=\; E|Z - X| \;=\; d_W(\mathcal{L}(X), \mathcal{L}(Z)).$$

It is worth indicating that an analogous argument allows us to show that the same bound (17) holds under the single condition $X + Y \succeq_{st} Z$. A priori, this result seems to be preferable, since the extra condition $Z - Y \geq 0$ is not required. One can see, however, that $X \succeq_{st} Z - Y$ does not imply $X + Y \succeq_{st} Z$ in general. As an example, choose $X = U$, $Y = U$ and $Z = n$ a.s., where n is any fixed positive integer and U is discrete uniform on the set $\{0, 1, \ldots, n\}$. Then, $X = U =_d n - U = Z - Y$ so that $X \succeq_{st} Z - Y$, but $X + Y = 2U$ is not $\succeq_{st}$ than $n = Z$.

3. A simple sufficient condition and examples

In practice, it may be difficult to check directly such conditions as stochastic ordering between W_α and W_β, as required by (5) and (10). It is thus useful to have available a simple sufficient condition which we may then apply.

Throughout this subsection, we assume that $\alpha = \beta$ and W_α and W_β have equal moments of order $t = 1, \ldots, s - 1$. That is, we assume

$$\text{condition } (A_s): \qquad E[\alpha_W (W + 1)^t] = E[\beta_W W^t], \qquad t = 0, \ldots, s - 1.$$

A well-known Karlin-Novikoff sufficient condition to guarantee the s-convex ordering in (10) under (A_s) is that our sequence $\{\alpha_{j-1}\mu_{j-1} - \beta_j \mu_j\}$ has at most s changes of sign.

Proposition 3. *Suppose that the condition (A_s) is satisfied and that the sequence $\{\alpha_{j-1}\mu_{j-1} - \beta_j \mu_j\}$ has at most s changes of sign. Then (11) holds.*

As a consequence, we obtain the following corollary, which extends Proposition A.1 of Barbour and Pugliese (2000) to birth-death processes.

Corollary 2. *Suppose that* $E\alpha_W = E\beta_W$. *If the sequence* $\{\alpha_{j-1}\mu_{j-1} - \beta_j\mu_j\}$ *is monotone, then* W_α *and* W_β *are stochastically ordered, so that the inequality (6) may be applied.*

We illustrate these results with the following examples.

Example 1. Our first example is motivated by Phillips and Weinberg (2000). Let W have a Bose-Einstein occupancy distribution. That is, given $m, d \geq 1$,

$$\mu_j = P(W = j) = \binom{d+m-j-2}{m-j}\binom{d+m-1}{m}^{-1}, \quad 0 \leq j \leq m.$$

We wish to approximate W by $\pi \sim \mathrm{Geom}(p)$ where $p = (d-1)/(d+m-1)$. Let $q = 1-p$. To obtain our geometric law, we choose $\alpha_j = q$ and $\beta_j = I_{(j>0)}$, $j \geq 0$ as birth and death rates.

Firstly, one can easily check that in this case, $E\alpha_W = E\beta_W$ and the sequence $\{q\mu_{j-1} - \mu_j\}$ is non-decreasing, so that $W_\alpha \succeq_{st} W_\beta$. Using Corollary 2, the bound (6) then becomes

$$|Eh(W) - Eh(\pi)| \;\leq\; p\,\|\Delta Sh\|_\infty\,|EW - E\pi|. \tag{18}$$

Moreover, it is known (see Peköz (1996, Section 2)) that the Stein operator S admits here the representation

$$Sh(j) = -\sum_{i=j}^{\infty}[h(i) - Eh(\pi)]\,q^{i-j}.$$

From this, we find that $\Delta Sh(k) = -\sum_{i=k}^{\infty}\Delta h(i)q^{i-k}$, which leads to the bound

$$\|\Delta Sh\|_\infty \leq\; p^{-1}\|\Delta h\|_\infty.$$

Inserting this bound in (18) yields the following.

Corollary 3. *With* W *and* π *as above,*

$$|Eh(W) - Eh(\pi)| \;\leq\; \|\Delta h\|_\infty\,\frac{m}{d(d-1)}.$$

In particular, $d_{TV}(\mathcal{L}(W), \mathcal{L}(\pi)) \leq m/d(d-1)$.

Example 2. Our next examples centre around approximation by so–called polynomial birth–death distributions, defined by Brown and Xia (2001) as the equilibrium distribution of a birth–death process with birth and death rates α_j and β_j which are polynomial in j. With such choices, we will write $\pi \sim \mathrm{PBD}(\alpha_j, \beta_j)$.

Suppose that W satisfies $\mu_j = (a+bj^{-1})\mu_{j-1}$ for some $a, b \in \mathbb{R}$. That is, W belongs to the Katz (or Panjer) family of distributions (see Johnson *et al.* (1992, Section 2.3.1)). It is well known that in this case W must have either either a binomial, Poisson or negative binomial distribution.

We fix some $l \geq 1$ and consider the approximation of W by the polynomial birth-death distribution $\pi \sim \mathrm{PBD}(\alpha, jQ_{l-1}(j))$. Here we have chosen a constant birth rate α and a death rate $\beta_j = jQ_{l-1}(j)$, where $Q_{l-1}(j)$ is a non–decreasing, monic polynomial in j of degree $l-1$. This gives us l parameters needed to specify the distribution of π. We choose these parameters in such a way that the condition (A_l) is satisfied.

With our choice of birth and death rates we have that

$$\alpha\mu_{j-1} - \beta_j\mu_j \;=\; \alpha\mu_{j-1} - jQ_{l-1}(j)(a+bj^{-1})\mu_{j-1} \;=\; \mu_{j-1}[\alpha - ajQ_{l-1}(j) - bQ_{l-1}(j)].$$

Noting that $\alpha - ajQ_{l-1}(j) - bQ_{l-1}(j)$ is a polynomial of degree l in j, and therefore has at most l real roots, we have that the sequence $\{\alpha_{j-1}\mu_{j-1} - \beta_j\mu_j\}$ has at most l changes of sign, so that either $W_\alpha \succeq_{l-cx} W_\beta$ or $W_\beta \succeq_{l-cx} W_\alpha$.

Theorem 2.10 of Brown and Xia (2001) gives us that

$$\sup\{\|\Delta Sh\|_\infty : h(j) = I_{(j\in B)}, B \subseteq \mathbb{Z}^+\} \leq \alpha^{-1}.$$

Hence, with $h(j) = I_{(j\in B)}$ for some $B \subseteq \mathbb{Z}^+$,

$$\|\Delta^l Sh\|_\infty \;\leq\; 2^{l-1}\|\Delta Sh\|_\infty \;\leq\; 2^{l-1}\alpha^{-1}.$$

From Corollary 1 we thus obtain Corollary 4.

Corollary 4. *With W and π as above,*

$$d_{TV}(\mathcal{L}(W), \mathcal{L}(\pi)) \;\leq\; 2^{l-1}\alpha^{-1}\left| E\left[\alpha\binom{W+1}{l} - WQ_{l-1}(W)\binom{W}{l}\right]\right|. \qquad (19)$$

For example, consider the case where $W \sim \mathrm{Bin}(n,p)$ and $\pi \sim \mathrm{PBD}(\alpha, \gamma j + j(j-1))$, so that $l = 2$. Choosing our constants α and γ according to the prescription above,

straightforward calculations give us that

$$\alpha = n(n-1)p(1-p), \qquad \text{and} \qquad \gamma = (n-1)(1-2p).$$

Furthermore,

$$E[W(W+1)] = np(np+2-p), \qquad E[W^2(W-1)] = n(n-1)p^2(np+2-2p), \qquad \text{and}$$
$$E[W^2(W-1)^2] = n(n-1)p^2(n^2p^2+4np-5np^2-8p+6p^2+2).$$

Evaluating the bound (19) then gives

Corollary 5. *Assume that $W \sim Bin(n,p)$ and $\pi \sim PBD(\alpha, \gamma j + j(j-1))$. Then,*

$$d_{TV}(\mathcal{L}(W), \mathcal{L}(\pi)) \leq 2p^2. \tag{20}$$

We note that (19) does not necessarily give a bound of the optimal order. In the case covered by (20), Theorem 3.1 of Brown and Xia (2001) gives a bound on total variation distance of order $O(p^2/\sqrt{\lambda})$, where $\lambda = E[W] = np$. This disparity is due to our rather crude use of the supremum norm in obtaining bounds such as (19). In Sections 5 and 6, we will consider more refined ways to bound the terms of our Stein equation in some particular cases when we have two parameters to choose in our approximating distribution π. Despite this disadvantage, we nevertheless note that (19) gives an explicit bound which may be applied in many contexts.

Example 3. Our final example of this section focuses on mixture distributions of the polynomial birth–death type. Suppose that $\pi \sim \text{PBD}(\alpha, \beta_j)$ and $W \sim \text{PBD}(\xi, \beta_j)$, for some constant birth rate α, polynomial death rate β_j and random variable ξ on $\mathbb{R}^+$. In this case we have that

$$\mu_j = \frac{E\left[\mu_0(\xi)\xi^j\right]}{\prod_{k=1}^{j}\beta_k}, \quad j \geq 0. \tag{21}$$

We choose α such that $\alpha = E\beta_W$, that is,

$$\alpha \;=\; E\sum_{j=0}^{\infty}\beta_j\mu_j \;=\; E\sum_{j=0}^{\infty}\xi\mu_{j-1} \;=\; E\xi.$$

Using (21), we obtain

$$\begin{aligned}\alpha\mu_j - \beta_{j+1}\mu_{j+1} \quad &= \quad E\left[\frac{\mu_0(\xi)\alpha^{j+1}}{\prod_{k=1}^{j}\beta_j}\left\{\left(\frac{\xi}{\alpha}\right)^j - \left(\frac{\xi}{\alpha}\right)^{j+1}\right\}\right] \\ &= \quad E\left[\frac{\alpha\mu_0(\xi)}{\mu_0(\alpha)}\left(1-\frac{\xi}{\alpha}\right)\pi_j\left(\frac{\xi}{\alpha}\right)^j\right].\end{aligned}$$

From this, we can see that the sequence $\{\alpha\mu_j - \beta_{j+1}\mu_{j+1}\}$ is monotone. Hence, Corollary 2 gives us the following.

Corollary 6. *With W and π as above,*

$$|Eh(W) - Eh(\pi)| \;\leq\; \|\Delta Sh\|_\infty \, |E[\alpha(W+1) - \beta_W W]| \,. \tag{22}$$

For example, if $\beta_j = j$ then $W \sim \mathrm{Po}(\xi)$ and we take $\pi \sim \mathrm{Po}(\lambda)$, where $\lambda = E\xi$. Using the well–known bound on the Stein operator S in this case, namely

$$\|\Delta Sh\|_\infty \leq \; \lambda^{-1}(1 - e^{-\lambda}) \, \|h\|_\infty, \tag{23}$$

evaluating (22) gives, after some straightforward calculation,

$$d_{TV}(\mathcal{L}(W), \mathrm{Po}(\lambda)) \;\leq\; \lambda^{-1}(1 - e^{-\lambda})\mathrm{Var}(\xi),$$

a bound that has also been obtained by Barbour *et al.* (1992, Theorem 1.C).

4. Poisson approximation for a sum of indicators

Throughout this section, the random variable W of interest is a sum of indicators:

$$W = X_1 + \cdots + X_n,$$

where the X_i are Bernoulli variables, possibly dependent, with

$$p_i = P(X_i = 1) = 1 - P(X_i = 0), \quad 1 \leq i \leq n.$$

Using Propositions 1 and 2, we are going to investigate the approximation of the sum W by a Poisson random variable $\pi \sim \mathrm{Po}(\lambda)$.

Recall that our Poisson variable is derived from (1) when $\alpha_j = \lambda$ and $\beta_j = j$, so that by (7),

$$W_\alpha = W + 1, \;\text{ and }\; P(W_\beta \in B) = \frac{E[WI_{(W\in B)}]}{EW}, \tag{24}$$

for any Borel set B. In the analysis, an important role will be played by the variables

$$W_i = W - X_i, \quad 1 \leq i \leq n.$$

4.1. Total dependence

Firstly, we consider the case where the indicators X_i are totally negatively dependent in the sense of Papadatos and Papathanasiou (2002). Let us recall that n random variables $X_i,\ 1 \leq i \leq n$, are totally negatively dependent (TND) if

$$\mathrm{Cov}[g_1(X_i), g_2(W_i)] \leq 0, \quad 1 \leq i \leq n, \tag{25}$$

for all non-decreasing functions g_1, g_2 such that the covariance exists.

Papadatos and Papathanasiou (2002, Theorem 3.1) show that the class of TND indicators includes the standard class of negatively related indicators. Stein's method for Poisson approximation of a sum of negatively related indicators is discussed by, for example, Barbour *et al.* (1992) and Erhardsson (2005). Recall that indicator random variables $X_1, \ldots, X_n$ are said to be negatively related if

$$E[g(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)|X_i = 1] \ \leq \ E[g(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)], \quad 1 \leq i \leq n, \tag{26}$$

for all non-decreasing functions $g : \{0,1\}^{n-1} \mapsto \{0,1\}$.

We wish to bound the total variation distance between $\mathcal{L}(W)$ and $\mathrm{Po}(\lambda)$. For that, we will apply Proposition 1. By (24), we have that, for any function $g : \mathbb{Z}^+ \mapsto \mathbb{R}$,

$$Eg(W_\alpha) = Eg(W+1), \ \text{ and } \ Eg(W_\beta) = \frac{E[Wg(W)]}{EW}.$$

Thus, to show that $W_\alpha \succeq_{st} W_\beta$, we must prove that if g is non-decreasing, then $EW Eg(W+1) \geq E[Wg(W)]$. In fact, this was established by Papadatos and Papathanasiou (2002, Lemma 3.1).

Using the bound (23) on the Stein operator in the Poisson case, (5) and (6) provide the following result.

Theorem 1. *If the indicators $\{X_i : 1 \leq i \leq n\}$ are TND, then $W_\alpha \succeq_{st} W_\beta$. If, in addition, $EW \geq \lambda$, then*

$$d_{TV}(\mathcal{L}(W), Po(\lambda)) \leq \frac{1-e^{-\lambda}}{\lambda}\left([\lambda+1]EW - E[W^2]\right).$$

Further results on, and examples of, TND indicator random variables can be found in Papadatos and Papathanasiou (2002).

Let us now consider the case where the indicators X_i are positively dependent in a certain sense. We adapt the definition (25) and say that n random variables $X_1 \dots, X_n$, are totally positively dependent (TPD) if

$$\mathrm{Cov}[g_1(X_i), g_2(W_i)] \geq 0, \quad 1 \leq i \leq n,$$

for all non-decreasing functions g_1, g_2 such that the covariance exists.

Association or positive relation is sufficient for TPD. This can be established analogously to the proof of Theorem 3.1 of Papadatos and Papathanasiou (2002). Recall that our indicator random variables are said to be positively related if (26) holds with the inequality reversed for all non-decreasing functions $g : \{0,1\}^{n-1} \mapsto \{0,1\}$. This standard property is used with Stein's method by, for example, Barbour *et al.* (1992) and Erhardsson (2005).

In the sequel, it is assumed that $EW = \lambda$. To get a bound for the total variation distance, we will apply Proposition 2, using the lemma stated below. To begin with, we introduce a random variable X_V, a mixing of our n indicators, in which the index V is a random variable of law

$$P(V = i) = \frac{EX_i}{\lambda}, \quad 1 \leq i \leq n. \tag{27}$$

Lemma 2. *If $EW = \lambda$ and the indicators $\{X_i : 1 \leq i \leq n\}$ are TPD, then*

$$W_\beta \succeq_{st} W_\alpha - X_V, \tag{28}$$

where $W_\alpha - X_V \geq 0$ a.s.

Proof. As seen in (24), $W_\alpha = W + 1$ and thus, $W_\alpha - X_V \geq 0$ a.s. Moreover, W_β has the so-called W-size-biased distribution: see, for example, Goldstein and Rinott (1996). W being a sum of indicators, it is then known that W_β admits the representation

$$W_\beta = \sum_{i \neq V} \hat{X}_i + 1, \tag{29}$$

where V is a random variable of law (27), and if $V = v$,

$$\hat{X}_i =_d (X_i | X_v = 1), \quad i \neq v.$$

Thus, by (29), the ordering (28) is equivalent to $\sum_{i\neq V}\hat{X}_i \succeq_{st} W - X_V$. To establish this, it is enough to prove that

$$\sum_{i\neq v}\hat{X}_i \succeq_{st} W - X_v, \quad 1 \leq v \leq n;$$

see Shaked and Shanthikumar (2007). Now, by (29) and the TPD assumption, we get, for any real $a \geq 0$,

$$\begin{aligned} P(\sum_{i\neq v}\hat{X}_i > a) &= P(\sum_{i\neq v} X_i > a | X_v = 1) \\ &\geq P(\sum_{i\neq v} X_i > a) \;=\; P(W - X_v > a), \end{aligned}$$

which is the desired result.

Thanks to Lemma 2, we may apply Proposition 2 with $s = p = 1$. Noting that by (27),

$$EX_V \;=\; \sum_{i=1}^n p_i\, P(V = i) \;=\; \frac{1}{\lambda}\sum_{i=1}^n p_i^2,$$

we then get the following result.

Theorem 2. *If $EW = \lambda$ and the indicators $\{X_i : 1 \leq i \leq n\}$ are TPD, then*

$$d_{TV}(\mathcal{L}(W), Po(\lambda)) \leq \frac{1 - e^{-\lambda}}{\lambda}\left\{E[W^2] + 2\sum_{i=1}^n p_i^2 - \lambda(\lambda + 1)\right\}.$$

This bound is obtained (and applied) by Barbour *et al.* (1992, Corollary 2.C.4) under the condition of positive relation. See also Erhardsson (2005).

4.2. Local dependence

Our goal in this part is to combine the previous s-convex ordering approach with a more flexible property of dependence. More precisely, we first introduce a concept of local dependence between a set of n indicators $X_1, \ldots, X_n$.

Let $\mathcal{F}_s$ be the class of all functions $g : \{0,1\}^{n-1} \mapsto \mathbb{R}^+$ that are non-decreasing and s-convex with $g(0) = 0$. We say that the n indicators $X_1 \ldots, X_n$, are (s,δ)-*locally negatively dependent* ((s,δ)-LND) if there exist n non-negative reals $\delta_1 \ldots, \delta_n$ (of sum > 0) such that

$$E[X_i g(W_i)] \;\leq\; \delta_i\, E[g(W_i)] \;\text{ for all functions } g \in \mathcal{F}_s, \quad 1 \leq i \leq n. \tag{30}$$

Similarly, $X_1 \ldots, X_n$, are said to be (s,δ)-locally positively dependent ((s,δ)-LPD) if

$$E[X_i g(W_i)] \;\geq\; \delta_i \, E[g(W_i)] \;\text{ for all functions } g \in \mathcal{F}_s, \quad 1 \leq i \leq n. \tag{31}$$

Let $\delta := \delta_1 + \ldots + \delta_n$, and denote

$$v_i = \delta^{-1}\delta_i, \quad 1 \leq i \leq n, \;\text{ and }\; p = EW/\delta \wedge \delta/EW.$$

We then adopt the notation v_p and X_V of Sections 1 and 4.1.

Lemma 3. *If the indicators $\{X_i : 1 \leq i \leq n\}$ are (s,δ)-LND, then*

$$W_\alpha \succeq_{s-cx} v_p W_\beta, \tag{32}$$

while if the indicators $\{X_i : 1 \leq i \leq n\}$ are (s,δ)-LPD, then

$$W_\beta \succeq_{s-cx} v_p(W_\alpha - X_V). \tag{33}$$

Proof. The method of proof is built on ideas in Barbour *at al.* (1992), Goldstein and Rinott (1996), Papadatos and Papathanasiou (2002) and Reinert (2005). Let g be any function belonging to $\mathcal{F}_s$. As a preliminary, we observe that $W \leq W_i + 1 \leq W + 1$ a.s. for each $i = 1, \ldots, n$.

Now, consider the case of (s,δ)-LND. Using (30) and the assumption that g is non-decreasing, we obtain that

$$\begin{aligned} E[Wg(W)] &= \sum_{i=1}^n E[X_i g(W)] \;=\; \sum_{i=1}^n E[X_i g(W_i+1)] \;\leq\; \sum_{i=1}^n \delta_i E[g(W_i+1)] \\ &\leq \sum_{i=1}^n \delta_i E[g(W+1)] \;=\; \delta E[g(W_\alpha)]. \end{aligned}$$

As $g(0) = 0$, and $EW/\delta \geq p \in (0,1]$, we find from (34) that

$$\begin{aligned} E[g(W_\alpha)] &\geq \frac{E[Wg(W)]}{EW}\,\frac{EW}{\delta} \\ &\geq pE[g(W_\beta)] \;=\; E[g(v_p W_\beta)], \end{aligned}$$

hence the ordering (32).

The case of (s,δ)-LPD is treated similarly. By (31) and since g is non-decreasing, we get

$$\begin{aligned} E[Wg(W)] &= \sum_{i=1}^n E[X_i g(W_i+1)] \;\geq\; \sum_{i=1}^n \delta_i E[g(W_i+1)] \\ &= \delta \sum_{i=1}^n P(V=i) E[g(W+1-X_i)] \;=\; \delta E[g(W_\alpha - X_V)]. \end{aligned} \tag{34}$$

As before, we then deduce from (34) that

$$\begin{aligned} E[g(W_\beta)] &= \frac{E[Wg(W)]}{EW} \;\geq\; E[g(W_\alpha - X_V)]\,\frac{\delta}{EW} \\ &\geq pE[g(W_\alpha - X_V)] \;=\; E[g(v_p(W_\alpha - X_V))], \end{aligned}$$

proving the ordering (33).

Combining Proposition 2 and Lemma 3 would then allow us to derive an upper bound for the total variation distance.

4.3. Approximate local dependence

Approximate local dependence is becoming a rather popular topic in probability. For works related to this idea, see for example Chen (1975), Barbour *et al.* (1992) and Chatterjee *et al.* (2005). We wish now to to derive an abstract Poisson approximation theorem by combining stochastic ordering with such an approach.

We say that the n indicators $X_1, \ldots, X_n$ are approximately locally negatively dependent (ALND) if there exist n non-negative reals $\delta_1, \ldots, \delta_n$ (of sum $\delta > 0$), and n random variables $Y_1, \ldots, Y_n$ on $\mathbb{Z}^+$ such that

$$E[X_i g(W_i - Y_i)] \;\leq\; \delta_i\, E[g(W_i - Y_i)], \quad 1 \leq i \leq n, \tag{35}$$

for all non-negative, non-decreasing functions g. Similarly, $X_1, \ldots, X_n$ are said to be approximately locally positively dependent (ALPD) if

$$E[X_i g(W_i - Y_i)] \;\geq\; \delta_i\, E[g(W_i - Y_i)], \quad 1 \leq i \leq n, \tag{36}$$

for all non-negative, non-decreasing functions g.

Define

$$\varepsilon = \sum_{i=1}^n E[X_i Y_i], \quad \text{and} \quad \varepsilon_* = \varepsilon + \sum_{i=1}^n \delta_i E[X_i + Y_i],$$

and let

$$c_\lambda = (\lambda+1)(1-e^{-\lambda})/\lambda + 2d_\lambda, \quad \text{with} \quad d_\lambda = 1 \wedge \sqrt{2/e\lambda}.$$

Theorem 3. *If $EW = \lambda$ and the indicators $\{X_i : 1 \leq i \leq n\}$ are ALND, then*

$$d_{TV}(\mathcal{L}(W), Po(\lambda)) \leq \frac{1-e^{-\lambda}}{\lambda}\,(|\,Var(W) - \lambda| + 2\varepsilon) + c_\lambda\, |\delta - \lambda|, \tag{37}$$

while if the indicators $\{X_i : 1 \leq i \leq n\}$ are ALPD, then

$$d_{TV}(\mathcal{L}(W), Po(\lambda)) \leq \frac{1-e^{-\lambda}}{\lambda}\,(|\,Var(W) - \lambda| + 2\varepsilon_*) + c_\lambda\, |\delta - \lambda|.$$

Before proving Theorem 3, we give an example of its application.

Example 4. We examine a variation of the classical birthday problem; see also Barbour *et al.* (1992). Suppose we independently colour $N \geq 2$ points with one of m colours, each colour being chosen equiprobably. Let Γ be the set of all subsets $i \subseteq \{1, \ldots, N\}$ of size 2. For $i \in \Gamma$, let Z_i be the indicator that the points indexed by i have the same colour. Moreover, suppose we choose uniformly r of the $|\Gamma| = \binom{N}{2}$ pairs of points, independently of the colourings chosen. For $i \in \Gamma$, we let $\xi_i = 0$ if the pair of points indexed by i is chosen, and otherwise set $\xi_i = 1$.

Set $W = \sum_{i \in \Gamma} Z_i \xi_i$. This counts the number of pairs of points with the same colour, excluding those r pairs of points we have chosen. In the case where $r = 0$, this corresponds to the classical birthday problem. A bound in the Poisson approximation of W in this case is given by Arratia *et al.* (1989, Example 2).

We observe that for all $i, j \in \Gamma$, $E[Z_i] = m^{-1}$ and $E[Z_i Z_j] = m^{-2}$. Furthermore,

$$E[\xi_i] = \frac{\binom{N}{2} - r}{\binom{N}{2}}, \qquad \text{and} \qquad E[\xi_i \xi_j] = \frac{\binom{N}{2} - r}{\binom{N}{2}} \left(\frac{\binom{N}{2} - r - 1}{\binom{N}{2} - 1} \right), \quad i \neq j.$$

Straightforward calculations then give

$$\lambda = E[W] = \frac{\binom{N}{2} - r}{m}, \qquad \text{and} \qquad \lambda - \mathrm{Var}(W) = \frac{\binom{N}{2} - r}{m^2}.$$

Now, we write $W_i = W - Z_i \xi_i$ and choose

$$Y_i = \sum_{j \neq i} Z_j \xi_j I_{(i \cap j \neq \emptyset)}, \quad \text{and} \quad \delta_i = E[Z_i \xi_i].$$

The condition (35) holds true with these choices. Indeed, $W_i - Y_i$ is independent of Z_i and the ξ_i are negatively related by construction. Thus, for all non–decreasing functions g, we have

$$E[Z_i \xi_i g(W_i - Y_i)] \;=\; E[Z_i \xi_i] \, E[g(W_i - Y_i) | \xi_i = 1] \;\leq\; E[Z_i \xi_i] \, E[g(W_i - Y_i)],$$

as required. We further see that

$$\begin{aligned} \varepsilon \;=\; \sum_{i \in \Gamma} E[Z_i \xi_i Y_i] \;&=\; \sum_{i \in \Gamma} \sum_{j \neq i} E[Z_i Z_j] \, E[\xi_i \xi_j] \, I_{(i \cap j \neq \emptyset)} \\ &=\; \frac{2(N-1) \left\{ \binom{N}{2} - r \right\} \left\{ \binom{N}{2} - r - 1 \right\}}{m^2 \left\{ \binom{N}{2} - 1 \right\}}. \end{aligned}$$

Evaluating (37) then gives the following bound.

Corollary 7. *With W as above,*

$$d_{TV}(\mathcal{L}(W), Po(\lambda)) \leq \frac{1-e^{-\lambda}}{m}\left\{1+4(N-1)\left(\frac{\binom{N}{2}-r-1}{\binom{N}{2}-1}\right)\right\}.$$

In the case $r=0$, a bound of the same order was established by Arratia *et al.* (1989, Example 2).

4.4. Proof of Theorem 3

(i) Consider the ALND case. We suppose first that f is any non-negative, non-decreasing function. Arguing as for Lemma 3, we have

$$\begin{aligned} E[Wf(W)] &= \sum_{i=1}^n E[X_i f(W)] = \sum_{i=1}^n E[X_i f(W_i+1)] \\ &= \sum_{i=1}^n E[X_i f(W_i - Y_i + 1)] + \sum_{i=1}^n E\{X_i[f(W_i+1) - f(W_i - Y_i + 1)]\}, \end{aligned}$$

which we denote by $T_1 + T_2$. We bound the sum T_2 by noting that

$$|f(x) - f(y)| \leq \|\Delta f\|_\infty \, |x-y|,$$

which yields

$$T_2 \leq \|\Delta f\|_\infty \sum_{i=1}^n E(X_i Y_i) = \|\Delta f\|_\infty \, \varepsilon.$$

For the sum T_1, by (35) and since f is non-decreasing, we get

$$T_1 \leq \sum_{i=1}^n \delta_i E[f(W_i - Y_i + 1)] \leq \sum_{i=1}^n \delta_i E[f(W+1)] = \delta E[f(W+1)].$$

Inserting these two bounds, we find that

$$\begin{aligned} E[\mathcal{A}f(W)] &= \lambda E[f(W+1)] - E[Wf(W)] \\ &\geq -(\delta - \lambda)E[f(W+1)] - \|\Delta f\|_\infty \, \varepsilon. \end{aligned} \tag{38}$$

To get an upper bound, we define a function $\tilde{f}$ on $\{0, 1, \ldots, n-1\}$ by

$$\tilde{f}(x) = \|f\|_\infty + \|\Delta f\|_\infty \, x - f(x). \tag{39}$$

Note that $\tilde{f}$ is, as f, a non-negative, non-decreasing function. By assumption, $EW = \lambda$ so that $E[\mathcal{A}1] = 0$; observe also that $E[\mathcal{A}W] = \lambda E[W+1] - E[W^2] = -[\mathrm{Var}(W) - \lambda]$.

Thus,

$$\begin{aligned} E[A\tilde{f}(W)] &= \|g\|_\infty \, E[A1] + \|\Delta f\|_\infty \, E[AW] - E[Af(W)] \\ &= -\|\Delta f\|_\infty \, [\mathrm{Var}(W) - \lambda] - E[Af(W)]. \end{aligned}$$

On the other hand, (38) is applicable to the function $\tilde{f}$, so that

$$E[A\tilde{f}(W)] \geq -(\delta - \lambda)E[\tilde{f}(W+1)] - \|\Delta\tilde{f}\|_\infty \, \varepsilon.$$

From these two formulas, we deduce that

$$E[Af(W)] \leq \|\Delta\tilde{f}\|_\infty \, \varepsilon + (\delta - \lambda)E[\tilde{f}(W+1)] + \|\Delta f\|_\infty \, |\mathrm{Var}(W) - \lambda|. \tag{40}$$

Now, let f be an arbitrary function. We start with the standard decomposition $f = f_+ - f_-$, where f_+ and f_- are non-negative, non-decreasing functions with, of course,

$$\|\Delta^j f_+\|_\infty \leq \|\Delta^j f\|_\infty, \;\text{ and }\; \|\Delta^j f_-\|_\infty \leq \|\Delta^j f\|_\infty, \quad j = 0, 1. \tag{41}$$

By (38) and (40), we obtain an upper bound

$$\begin{aligned} E[Af(W)] &= E[Af_+(W)] - E[Af_-(W)] \\ &\leq \|\Delta\tilde{f}_+\|_\infty \, \varepsilon + (\delta - \lambda)E[\tilde{f}_+(W+1)] + \|\Delta f_+\|_\infty \, |\mathrm{Var}(W) - \lambda| \\ &\qquad + (\delta - \lambda)E[f_-(W+1)] + \|\Delta f_-\|_\infty \, \varepsilon \\ &= \|\Delta f_+\|_\infty \, |\mathrm{Var}(W) - \lambda| + (\|\Delta\tilde{f}_+\|_\infty + \|\Delta f_-\|_\infty) \, \varepsilon \\ &\qquad + (\delta - \lambda) \, \{\|f_+\|_\infty + \|\Delta f_+\|_\infty \, (\lambda + 1) - E[f(W+1)]\}, \end{aligned}$$

using (39) and $EW = \lambda$ for the last equality. By a similar method, we find as a lower bound

$$\begin{aligned} E[Af(W)] &\geq -(\delta - \lambda)E[f_+(W+1)] - \|\Delta f_+\|_\infty \, \varepsilon \\ &\qquad - \|\Delta\tilde{f}_-\|_\infty \, \varepsilon - (\delta - \lambda)E[\tilde{f}_-(W+1)] - \|\Delta f_-\|_\infty \, |\mathrm{Var}(W) - \lambda| \\ &= -\|\Delta f_-\|_\infty \, |\mathrm{Var}(W) - \lambda| - (\|\Delta f_+\|_\infty + \|\Delta\tilde{f}_-\|_\infty) \, \varepsilon \\ &\qquad - (\delta - \lambda) \, \{\|f_-\|_\infty + \|\Delta f_-\|_\infty \, (\lambda + 1) + E[f(W+1)]\}. \end{aligned}$$

By (41) and since $\|\Delta\tilde{f}\|_\infty \leq \|\Delta f\|_\infty$, combining the two previous bounds then yields

$$|E[Af(W)]| \leq \|\Delta f\|_\infty \, (|\mathrm{Var}(W) - \lambda| + 2\varepsilon) + |\delta - \lambda| \, [2\|f\|_\infty + \|\Delta f\|_\infty \, (\lambda + 1)]. \tag{42}$$

With $f = Sh$, it now suffices to apply in (42) the standard bounds

$$\|\Delta Sh\|_\infty \le \lambda^{-1}(1-e^{-\lambda})\|h\|_\infty, \text{ and } \|Sh\|_\infty \le d_\lambda \|h\|_\infty,$$

which gives (37).

(ii) The ALPD case is dealt with analogously. For f non-negative, non-decreasing, we first write that

$$\begin{aligned} E[Wf(W)] &= \sum_{i=1}^n E[X_i f(W_i - Y_i + 1)] + \sum_{i=1}^n E[X_i\{f(W_i+1) - f(W_i - Y_i + 1)\}] \\ &\ge \sum_{i=1}^n E[X_i f(W_i - Y_i + 1)] - \|\Delta f\|_\infty\, \varepsilon. \end{aligned}$$

By (36), we then get that

$$\begin{aligned} E[Wf(W)] &\ge \sum_{i=1}^n \delta_i E[f(W_i - Y_i + 1)] - \|\Delta f\|_\infty\, \varepsilon \\ &= \delta E[f(W+1)] - \sum_{i=1}^n \delta_i E[f(W+1) - f(W_i - Y_i + 1)] - \|\Delta f\|_\infty\, \varepsilon \\ &\ge \delta E[f(W+1)] - \|\Delta f\|_\infty \sum_{i=1}^n \delta_i E(X_i + Y_i) - \|\Delta f\|_\infty\, \varepsilon \\ &= \delta E[f(W+1)] - \|\Delta f\|_\infty\, \varepsilon_*. \end{aligned}$$

Overall, we find that

$$E[Af(W)] = \lambda E[f(W+1)] - E[Wf(W)] \ge -(\delta - \lambda)E[f(W+1)] + \|\Delta f\|_\infty\, \varepsilon_*.$$

The rest of the proof follows as in the ALND case.

5. Translated Poisson approximation

We assume, as in Section 4, that $W = X_1 + \cdots + X_n$ is a sum of (possibly dependent) indicator random variables, with $p_i = P(X_i = 1)$. Denote

$$\lambda_k = \sum_{i=1}^n p_i^k, \quad \lambda = \lambda_1 = E[W], \quad \text{and} \quad \sigma^2 = \mathrm{Var}(W).$$

We are going to discuss the approximation of W by a translated Poisson distribution.

5.1. Main results

A random variable Z has a translated Poisson distribution $\mathrm{TP}(\lambda, \sigma^2)$ if Z is distributed as $Z' + \rho$, where $Z' \sim \mathrm{Po}(\sigma^2 + \gamma)$ with

$$\rho = \lambda - \sigma^2 - \gamma, \quad \text{and} \quad \gamma = \langle \lambda - \sigma^2 \rangle \in [0, 1),$$

$\langle x \rangle = x - \lfloor x \rfloor$ denoting the fractional part of x.

We note that $E[Z] = \lambda$ and $\sigma^2 \leq \mathrm{Var}(Z) = \sigma^2 + \gamma < \sigma^2 + 1$, so that our approximating translated Poisson distribution has a mean equal to, and variance close to, that of W. We would thus expect a closer approximation than could be obtained by simply using the one–parameter Poisson distribution. The variances of W and Z cannot necessarily be made to match exactly, as we must shift our Poisson distribution by an integer. However, the error term arising from this mismatch does not adversely affect the order of the bounds we obtain, as we shall see below.

The following results give us bounds in translated Poisson approximation for W under some stochastic ordering assumptions. We defer the proofs of Theorems 4 and 5 until Section 5.3, giving first some examples of their application, in Section 5.2.

Our bounds demonstrate convergence to a translated Poisson distribution if $\sigma \to \infty$ as $n \to \infty$. Bounds on the total variation distance between $\mathcal{L}(W)$ and a translated Poisson random variable may still be found if this is not the case, but require a different analysis of the error terms. For example, in proving Theorems 4 and 5, we write $P(W - \rho < 0) \leq \sigma^{-2}$. This error term may be reduced, or even omitted altogether depending on the problem at hand, with a more careful analysis. This could give us good bounds in cases where $\sigma \to \sigma_\infty < \infty$ as $n \to \infty$.

In the sequel, we let W^s be a random variable having the W–size–biased distribution, and v_q be an indicator random variable, independent of all else, with $P(v_q = 1) = q$. As before, we write $W_i = W - X_i$, $1 \leq i \leq n$, and for any random index V we let $W_V = W - X_V$.

Theorem 4. *Suppose that $X_1, \ldots, X_n$ are negatively related, and there is $q \in [0, 1]$ and $l \in \mathbb{Z}^+$ such that*

$$(W + 1 | X_k = 0) \preceq_{st} (W + l + v_q | X_k = 1), \quad 1 \leq k \leq n. \tag{43}$$

Then,

$$d_{TV}(\mathcal{L}(W), TP(\lambda, \sigma^2)) \leq \frac{2}{\sigma^2} + \frac{\lambda_2 + (l+q)(\lambda - \lambda_2)}{\lambda\sigma} + \frac{l(l+2q-1)(\lambda-\lambda_2)}{\sigma^2} d_{TV}(\mathcal{L}(W^s), \mathcal{L}(W^s+1)). \quad (44)$$

Theorem 5. *Suppose that $X_1, \dots, X_n$ are positively related, and there is $q \in [0,1]$ and $l \in \mathbb{Z}^+$ such that*

$$(W+1|X_k = 0) \succeq_{st} (W - l - v_q|X_k = 1), \quad 1 \leq k \leq n. \quad (45)$$

Then,

$$d_{TV}(\mathcal{L}(W), TP(\lambda, \sigma^2)) \leq \frac{2}{\sigma^2} + \frac{\lambda_2 + (l+q)(\lambda - \lambda_2)}{\lambda\sigma} + \frac{(l+1)(l+2q)(\lambda-\lambda_2)}{\sigma^2} d_{TV}(\mathcal{L}(W^s), \mathcal{L}(W^s+1)). \quad (46)$$

Consider the stochastic ordering assumptions (43) and (45). We note that the choice of l and q is not unique, in that choosing $l = m$, $q = 1$ gives the same assumption as choosing $l = m+1$, $q = 0$. It is easily checked, however, that each of these choices gives rise to the same bounds in (44) and (46). In the examples below, we will verify the validity of such stochastic orderings by using an appropriate coupling argument.

5.2. Applications

Example 5. Suppose that $X_1, \dots, X_n$ are independent. Thus, they are also negatively related. Moreover, the condition (43) is true for $q = l = 0$. Therefore, (44) is applicable and yields the following.

Corollary 8. *With W as above,*

$$d_{TV}(\mathcal{L}(W), TP(\lambda, \sigma^2)) \leq \frac{\lambda_2}{\lambda\sigma} + \frac{2}{\sigma^2}.$$

This bound is of the order we would expect: see also Čekanavičius and Vaĭtkus (2001).

Example 6. Suppose that m balls are placed into N urns, in such a way that no urn contains more than one ball and all arrangements are equally likely. Let W be the number of balls in the first n urns. Thus, W has a hypergeometric distribution with

$$\lambda = \frac{mn}{N}, \quad \text{and} \quad \sigma^2 = \frac{mn(N-m)(N-n)}{(N-1)N^2}.$$

We set X_i to be the indicator that the ith urn contains a ball, so that $W = X_1 + \cdots + X_n$. By construction, these indicators are negatively related. The condition (43) holds for $q = 1$ and $l = 0$. To see this, we construct $(W + 1|X_k = 0)$ by considering the N urns and excluding the kth. Distribute the m balls in these $N - 1$ urns, such that all arrangements are equally likely, and count the number of the first n urns that are occupied. Adding one to this count gives us our random variable. We then choose (uniformly and independently of what has gone before) one of the occupied urns. Take the ball from this urn and place it in urn k. This gives us $(W + 1|X_k = 1)$. If the ball chosen is from one of the first n urns, the number of occupied urns is the same as before. Otherwise, we have increased the number of occupied urns within the first n. Evaluating the bound (44) then gives Corollary 9.

Corollary 9. *For W having our hypergeometric distribution,*

$$d_{TV}(\mathcal{L}(W), TP(\lambda, \sigma^2)) \;\leq\; \frac{1}{\sigma} + \frac{2}{\sigma^2} \;=\; \sqrt{\frac{N^2(N-1)}{mn(N-m)(N-n)}} + \frac{2N^2(N-1)}{mn(N-m)(N-n)}.$$

Röllin (2007, Section 4.1) has considered translated Poisson approximation for the hypergeometric distribution, and shows that if $m = O(n)$ and $N = O(n)$, then one gets a bound in total variation distance of order $O(1/\sqrt{n})$. This order is also reflected in our result.

Example 7. Suppose $\xi_1, \ldots, \xi_n$ are i.i.d. Bernoulli random variables with

$$p = P(\xi_i = 1) = 1 - P(\xi_i = 0), \quad 1 \leq i \leq n.$$

Fix an integer $k \geq 2$, and define

$$X_i = \xi_i \xi_{i+1} \cdots \xi_{i+k-1}, \quad \text{and} \quad W = \sum_{i=1}^{n} X_i,$$

in which, to avoid edge effects, all indices are treated modulo n. Thus, W counts the number of k–runs in our Bernoulli trials. Observe that

$$\lambda = np^k, \quad \lambda_2 = np^{2k}, \quad \text{and} \quad \sigma^2 = \frac{np^k}{1-p}\,(1 + p - p^k[2 + (2k-1)(1-p)]).$$

Translated Poisson approximation for k–runs was treated by Röllin (2005, Section 3.2), who gives a bound in total variation distance of the form $K/\sqrt{n}$, for some constant $K = K(k, p)$ independent of n. Barbour and Xia (1999, Section 5) also give a bound

of this order for 2–runs. We shall use our Theorem 5 to give an explicit bound with this same order.

It is easily seen that the variables $X_1, \dots, X_n$ are positively related. The condition (45) holds by choosing $q = 1$ and $l = 2k - 3$. To see that, consider the following construction. Given the Bernoulli random variables $\xi_1, \dots, \xi_n$, fix some $m \le n$ and set $\xi_m = \xi_{m+1} = \cdots = \xi_{m+k-1} = 1$, while the others remain independent Bernoulli random variables with parameter p. Counting the number of k–runs in these n Bernoulli trials gives us $(W|X_m = 1)$. Suppose now we resample the random variables $\xi_m, \dots, \xi_{m+k-1}$, conditional on at least one of these being zero. Counting the number of k–runs now gives us $(W|X_m = 0)$. In this resampling procedure, one can remove at most $2k-1$ of the k–runs that were originally present. Thus, our construction implies that $(W|X_m = 0) + 2k - 1 \ge (W|X_m = 1)$, or, equivalently, $(W + 1|X_m = 0) \ge (W - 2k + 2|X_m = 1)$, hence the announced values of q and l.

Following the work of Section 4, to construct W^s we choose an index V uniformly from $\{1, \dots, n\}$, and set $\xi_V = \xi_{V+1} = \cdots = \xi_{V+k-1} = 1$, while the other ξ_i remain independent Bernoulli random variables with parameter p. Lemma 2.1 of Wang and Xia (2008) thus gives us that

$$d_{TV}(\mathcal{L}(W^s), \mathcal{L}(W^s + 1)) \le 1 \wedge \frac{2.3}{\sqrt{(n-k-1)p^k(1-p)^3}}.$$

Using this, Theorem 5 yields the following.

Corollary 10. *Let W count the number of k–runs in n independent Bernoulli trials, each with success probability p. Then,*

$$d_{TV}(\mathcal{L}(W), TP(\lambda, \sigma^2)) \le \frac{2}{\sigma^2} + \frac{p^k + (2k-2)(1-p^k)}{\sigma} + \frac{(2k-2)(2k-1)np^k(1-p^k)}{\sigma^2}\left(1 \wedge \frac{2.3}{\sqrt{(n-k-1)p^k(1-p)^3}}\right). \quad (47)$$

Our bound (47) has the same order as that of Röllin (2005, Theorem 5) and Barbour and Xia (1999, Theorem 5.2) (this latter result applying only to the 2–runs case). Numerical comparison of the bounds shows that ours generally performs well compared to these other bounds, often giving a better result. Table 1 gives some illustrations, with values for comparison taken from Röllin (2005).

Table 1: Numerical comparisons for 2–runs. Upper bounds on total variation distance from (a) our result (47), (b) Röllin (2005) and (c) Barbour and Xia (1999). Missing values are due to restrictions on choice of parameters.

		$p = 0.10$	$p = 0.25$	$p = 0.50$	$p = 0.75$	$p = 0.90$
	(a)	0.1553	0.0675	0.0500	0.0814	0.2512
$n = 10^6$	(b)	0.4463	0.2334	0.1747	0.5528	> 1
	(c)	0.0304	–	0.1251	0.6014	–
	(a)	0.0155	0.0067	0.0050	0.0081	0.0251
$n = 10^8$	(b)	0.0445	0.0233	0.0175	0.0553	0.2554
	(c)	0.0030	–	0.0125	0.0601	–
	(a)	0.0016	0.0007	0.0005	0.0008	0.0025
$n = 10^{10}$	(b)	0.0045	0.0023	0.0017	0.0055	0.0255
	(c)	0.0003	–	0.0013	0.0060	–

5.3. Proof of Theorems 4 and 5

Our proof is based on that of Propositions 1 and 2, using the characterising operator for the Poisson distribution. We find representations of our Stein equation in conjunction with which our dependence and stochastic ordering assumptions may be applied.

Throughout this section we let $f = Sh$ be the solution to the Stein equation (2) with the choices $\alpha_j = \sigma^2 + \gamma$ and $\beta_j = j$, corresponding to the Poisson distribution with mean $\sigma^2 + \gamma$. We suppose the test function h has the form $h(j) = I_{(j \in B)}$ for some $B \subseteq \mathbb{Z}^+$. We write $g_B(j) = f(j - \rho)$. We note that g_B depends on the choice of set B, though for notational convenience we will often write simply g for g_B. We note further that bounds on the supremum norm of f also apply to g, so that in particular $\|\Delta g_B\|_\infty \leq \sigma^{-2}$ for each $B \subseteq \mathbb{Z}^+$.

Following Röllin (2007, Section 3), we obtain from the Stein equation that

$$d_{TV}(\mathcal{L}(W), TP(\lambda, \sigma^2)) \leq \sup_{B \subseteq \mathbb{Z}^+} |E[(\sigma^2 + \gamma) g_B(W+1) - (W - \rho) g_B(W)]| + P(W - \rho < 0). \tag{48}$$

One may bound $P(W - \rho < 0) \leq \sigma^{-2}$ using Chebyshev's inequality. So, we now concentrate on the first term on the right–hand side of (48). Throughout our proof,

we will make use of the following equalities in distribution:

$$(W|X_V = 1) =_d W^s, \qquad \text{and} \qquad (W_V|X_V = 0) =_d (W|X_V = 0). \tag{49}$$

Step (1). For this part of the proof, we will consider separately the cases where $\sigma^2 \le \lambda$ and $\sigma^2 \ge \lambda$. We begin by assuming $\sigma^2 \le \lambda$, so that $\rho \ge 0$. Recall that

$$E[Wg(W)] = \lambda E[g(W^s)]. \tag{50}$$

Using (50), we can then write that

$$E[(\sigma^2 + \gamma)g(W + 1) - (W - \rho)g(W)] = \lambda E[g(\widetilde{W}) - g(W^s)], \tag{51}$$

where

$$P(\widetilde{W} = j) = \lambda^{-1} \left\{(\sigma^2 + \gamma)P(W + 1 = j) + \rho P(W = j)\right\}, \quad j \ge 0.$$

That is, $\widetilde{W} = W + v_r$ where v_r is a Bernoulli variable with success probability $r = \lambda^{-1}(\sigma^2 + \gamma)$. Note that $r \le 1$ by assumption. We rewrite (51) as

$$\lambda E[g(\widetilde{W}) - g(W^s)] = \lambda E[g(\widetilde{W}) - g(\overline{W})] + \lambda E[g(\overline{W}) - g(W^s)], \tag{52}$$

by defining $\overline{W} = W_V + 1$, where V is a random index chosen according to (27). For the first term in (52) we note that, by conditioning on v_r,

$$\lambda Eg(\widetilde{W}) \;=\; \lambda Eg(W + v_r) \;=\; (\sigma^2 + \gamma)E\Delta g(W) + \lambda Eg(W). \tag{53}$$

Furthermore, by conditioning on X_V and using the equalities (49),

$$\lambda Eg(\overline{W}) \;=\; \lambda Eg(W_V + 1) \;=\; \lambda_2 Eg(W^s) + (\lambda - \lambda_2)E[g(W)|X_V = 0], \tag{54}$$

since $P(X_V = 1) = \lambda^{-1}\lambda_2$. Again by considering conditioning on X_V and using (49), we have that

$$(\lambda - \lambda_2)E[g(W)|X_V = 0] \;=\; \lambda Eg(W + 1) - \lambda_2 Eg(W^s + 1). \tag{55}$$

Combining (53), (54) and (55) we obtain the following.

$$\begin{aligned}\lambda E[g(\widetilde{W}) - g(\overline{W})] \;&=\; (\sigma^2 + \gamma - \lambda)E\Delta g(W) + \lambda_2 E\Delta g(W^s) \\ &=\; \lambda_2 E[\Delta g(W^s) - \Delta g(W)] + \gamma E\Delta g(W) \\ &\qquad + (\sigma^2 - \lambda + \lambda_2)E\Delta g(W).\end{aligned} \tag{56}$$

Now consider the second term of (52). Let us combine it with the final term of (56). Since

$$E[\overline{W} - W^s] = -\lambda^{-1}(\sigma^2 - \lambda + \lambda_2),$$

and proceeding as we did in deriving (3), we get that

$$\begin{aligned}\lambda E[g(\overline{W}) - g(W^s)] + (\sigma^2 - \lambda + \lambda_2) E\Delta g(W) \\ = \lambda E \sum_{j=0}^{\infty} (\Delta g(j) - \Delta g(W)) \big[P(\overline{W} > j) - P(W^s > j)\big]. \quad (57)\end{aligned}$$

Using the definition of $\overline{W}$, conditioning on X_V and employing (49), we have that

$$\begin{aligned}\lambda\big[P(\overline{W} > j) - P(W^s > j)\big] \\ = (\lambda - \lambda_2)\big[P(W_V + 1 > j|X_V = 0) - P(W_V + 1 > j|X_V = 1)\big]. \quad (58)\end{aligned}$$

Hence, the right–hand side of (57) becomes

$$(\lambda - \lambda_2) E \sum_{j=0}^{\infty} (\Delta g(j) - \Delta g(W)) \big[P(W_V + 1 > j|X_V = 0) - P(W_V + 1 > j|X_V = 1)\big]. \quad (59)$$

Let us now insert the representations (56) and (59) into (51) and then (48). We obtain

$$\begin{aligned}d_{TV}(\mathcal{L}(W), \mathrm{TP}(\lambda, \sigma^2)) \le (\lambda - \lambda_2) \sup_{B \subseteq \mathbb{Z}^+} \big\{\Lambda_B\big\} + \lambda_2 \sup_{B \subseteq \mathbb{Z}^+} \big|E[\Delta g_B(W^s) - \Delta g_B(W)]\big| \\ + \gamma \sup_{B \subseteq \mathbb{Z}^+} \big|E\Delta g_B(W)\big| + P(W - \rho < 0),\end{aligned}$$

where

$$\Lambda_B = E \sum_{j=0}^{\infty} |\Delta g_B(j) - \Delta g_B(W)||P(W_V + 1 > j|X_V = 0) - P(W_V + 1 > j|X_V = 1)|.$$

Recalling that $P(W - \rho < 0) \le \sigma^{-2}$, $\gamma \le 1$ and $\|\Delta g_B\|_\infty \le \sigma^{-2}$, we have that

$$\gamma \sup_{B \subseteq \mathbb{Z}^+} \big|E\Delta g_B(W)\big| + P(W - \rho < 0) \ \le\ 2\sigma^{-2}.$$

Furthermore, the random variable W^s having the W–size–biased distribution satisfies

$$P(W^s = j) \ =\ \lambda^{-1} j P(W = j), \quad 0 \le j \le n,$$

and so,

$$2d_{TV}(\mathcal{L}(W), \mathcal{L}(W^s)) \ =\ \sum_{j=0}^{\infty} |P(W = j) - P(W^s = j)| \ =\ E|1 - \lambda^{-1} W| \ \le\ \lambda^{-1}\sigma. \quad (60)$$

We thus have that

$$\lambda_2\big|E[\Delta g_B(W^s)-\Delta g_B(W)]\big| \;\leq\; 2\lambda_2\|\Delta g_B\|_\infty d_{TV}(\mathcal{L}(W),\mathcal{L}(W^s)) \;\leq\; \frac{\lambda_2}{\lambda\sigma}.$$

Combining the above bounds, we obtain

$$d_{TV}(\mathcal{L}(W),\mathrm{TP}(\lambda,\sigma^2)) \leq (\lambda-\lambda_2)\sup_{B\subseteq\mathbb{Z}^+}\big\{\Lambda_B\big\}+\frac{\lambda_2}{\lambda\sigma}+\frac{2}{\sigma^2}. \tag{61}$$

In the second step of the proof, we consider how Λ_B may be bounded. Before doing this, we show that if $\sigma^2\geq\lambda$ then the bound (61) continues to hold.

Consider now the case where $\sigma^2\geq\lambda$, so that $\rho\leq 0$. We will use an analogous argument to show that the bound (61) continues to hold. In place of (52), we this time write

$$\begin{aligned}E[(\sigma^2+\gamma)g(W+1)-(W-\rho)g(W)] &= (\sigma^2+\gamma)E[g(W+1)-g(\widehat{W})]\\ &\quad+(\sigma^2+\gamma)E[g(\widehat{W})-g(W^\star)], \end{aligned}\tag{62}$$

where $\widehat{W}=W+v_t(1-X_V)$, $W^\star=v_tW^s+(1-v_t)W$ and $t=\lambda(\sigma^2+\gamma)^{-1}$. Consider the first term on the right–hand side of (62). For this term, we argue as we did to derive (56). Conditioning on v_t and X_V and employing the equalities (49), we find, as for (56), that

$$\begin{aligned}&(\sigma^2+\gamma)E[g(W+1)-g(\widehat{W})]\\ &\qquad=\lambda_2E[\Delta g(W^s)-\Delta g(W)]+\gamma E\Delta g(W)+(\sigma^2-\lambda+\lambda_2)E\Delta g(W).\end{aligned}$$

As we have that

$$E[\widehat{W}-W^\star]=-(\sigma^2+\gamma)^{-1}(\sigma^2-\lambda+\lambda_2),$$

we then write

$$\begin{aligned}&(\sigma^2+\gamma)E[g(\widehat{W})-g(W^\star)]+(\sigma^2-\lambda+\lambda_2)E\Delta g(W)\\ &\qquad=(\sigma^2+\gamma)E\sum_{j=0}^{\infty}\left(\Delta g(j)-\Delta g(W)\right)[P(\widehat{W}>j)-P(W^\star>j)].\end{aligned}\tag{63}$$

Using the definitions of $\widehat{W}$ and $W^\star$, and conditioning on v_t, we find that

$$P(\widehat{W}>j)-P(W^\star>j)=t\left[P(\overline{W}>j)-P(W^s>j)\right].$$

Comparing this with (57), recalling the definition of t and using (58), we find that (59) also gives us a representation of (63). Continuing the argument as before, the bound (61) holds too in the present case.

Step (2). In this part of the proof, we bound Λ_B, and thus obtain the bounds of our theorems. In doing so, we will use our stochastic ordering and dependence assumptions. The cases where $X_1, \ldots, X_n$ are positively and negatively related will be discussed separately. In the positive related case, the argument of Lemma 2 shows that

$$P(W_V + 1 > j | X_V = 0) - P(W_V + 1 > j | X_V = 1) \leq 0, \quad j \geq 0.$$

Noting that $(W_V + 1 | X_V = 1) =_d W^s$, we fix some $l \in \mathbb{Z}^+$ and write

$$\begin{aligned} & P(W_V + 1 > j | X_V = 1) - P(W_V + 1 > j | X_V = 0) \\ & = P(W_V + 1 > j + l | X_V = 1) - P(W_V + 1 > j | X_V = 0) + \sum_{i=1}^{l} P(W^s = j + i). \end{aligned} \tag{64}$$

Suppose now that there is some $q \in [0, 1]$ such that for each $j \geq 0$

$$\begin{aligned} P(W_V + 1 > j + l | X_V = 1) - P(W_V + 1 > j | X_V = 0) & \\ \leq q\, P(W_V = j + l | X_V = 1) & \quad (65) \\ = q\, P(W^s = j + l + 1). & \quad (66) \end{aligned}$$

We will show presently that this is implied by the stochastic ordering assumption (45). Using (64) and (66), we find that

$$\begin{aligned} \Lambda_B & \leq qE|\Delta g_B(W^s - l - 1) - \Delta g_B(W)| + \sum_{i=1}^{l} E|\Delta g_B(W^s - i) - \Delta g_B(W)| \\ & \leq 2q\|\Delta g_B\|_\infty d_{TV}(\mathcal{L}(W), \mathcal{L}(W^s - l - 1)) + 2\|\Delta g_B\|_\infty \sum_{i=1}^{l} d_{TV}(\mathcal{L}(W), \mathcal{L}(W^s - i)). \end{aligned} \tag{67}$$

Using our bound on $\|\Delta g_B\|_\infty$ and the triangle inequality for total variation distance, the first term of (67) is bounded by

$$\begin{aligned} & 2q\sigma^{-2}\big\{d_{TV}(\mathcal{L}(W), \mathcal{L}(W^s)) + (l+1)d_{TV}(\mathcal{L}(W^s), \mathcal{L}(W^s + 1))\big\} \\ & \leq 2q\sigma^{-2}\left\{\frac{\sigma}{2\lambda} + (l+1)d_{TV}(\mathcal{L}(W^s), \mathcal{L}(W^s + 1))\right\}, \end{aligned} \tag{68}$$

where this last inequality uses (60). Similarly, the second term of (67) may be bounded by

$$2\sigma^{-2}\sum_{i=1}^{l}\left\{d_{TV}(\mathcal{L}(W),\mathcal{L}(W^s))+i\,d_{TV}(\mathcal{L}(W^s),\mathcal{L}(W^s+1))\right\}$$
$$\leq \sigma^{-2}\left\{\frac{l\sigma}{\lambda}+l(l+1)d_{TV}(\mathcal{L}(W^s),\mathcal{L}(W^s+1))\right\}. \quad (69)$$

Combining (67), (68) and (69) with the bound (61) yields the desired inequality (46).

So, the proof of Theorem 5 is completed upon showing that the stochastic ordering condition (45) implies the inequality (65). Writing

$$P(W_V=j+l|X_V=1)=P(W_V+1>j+l|X_V=1)-P(W_V>j+l|X_V=1),$$

for $0\leq j\leq n$, it can be seen that (65) is equivalent to

$$P(W_V+1>j|X_V=0)\geq(1-q)P(W_V+1-l>j|X_V=1)+qP(W_V-l>j|X_V=1),$$

for $j\geq 0$. This, in turn, is equivalent to the stochastic ordering

$$(W+1|X_V=0)\succeq_{st}(1-v_q)(W-l|X_V=1)+v_q(W-l-1|X_V=1), \quad (70)$$

which can be seen using (49). Some rearranging shows that the stochastic ordering assumption (45) implies the stochastic ordering (70), hence the result of Theorem 5.

We turn our attention now to the case of negative relation, and complete the proof of Theorem 4. When $X_1,\ldots,X_n$ are negatively related, one can use a similar argument to the above. We have here that

$$P(W_V+1>j|X_V=0)-P(W_V+1>j|X_V=1)\geq 0, \quad 0\leq j\leq n.$$

Analogously to the positively related case, we write, for some fixed $l\in\mathbb{Z}^+$,

$$P(W_V+1>j|X_V=0)-P(W_V+1>j|X_V=1)$$
$$=P(W_V+1>j|X_V=0)-P(W_V+1>j-l|X_V=1)+\sum_{i=0}^{l-1}P(W^s=j-i).$$

This time, we suppose that there is $q\in[0,1]$ such that

$$P(W_V+1>j|X_V=0)-P(W_V+1>j|X_V=1)\leq qP(W_V+1+l=j|X_V=1). \quad (71)$$

Following a similar argument to that used in the case of positive relation, we find that

$$\Lambda_B \leq \frac{l+q}{\lambda\sigma} + \frac{l(l+2q-1)}{\sigma^2}\, d_{TV}(\mathcal{L}(W^s), \mathcal{L}(W^s+1)).$$

Combining this with (61) gives us the desired inequality (44). It remains to show that the stochastic ordering assumption (43) implies the inequality (71), which can be done as above.

6. Another abstract approximation theorem

Our aim hereafter is to consider an alternative approximation theorem which may be found within the present framework. For concreteness, we suppose that the birth rates α_j and death rates β_j are such that the random variable π has two parameters available to choose. This will be the case in the application presented afterwards.

Let us return to the basic representation (12). To choose the two parameters of π, it seems natural, in our context, to consider $s = 2$ and introduce the two conditions $\alpha = \beta$ and $EW_\alpha = EW_\beta$ (i.e., $E[\alpha_W(W+1)] = E[\beta_W W]$). With these choices, the representation (12) becomes

$$Eh(W) - Eh(\pi) = \alpha \sum_{i=0}^{\infty} \Delta^2 f(i)\, E[(W_\alpha - i - 1)_+ - (W_\beta - i - 1)_+]. \tag{72}$$

Moreover, suppose that one can construct W_α and W_β on the same probability space in such a way that $W_\beta = W_\alpha + Y$ for some random variable Y which takes values in the set $\{-1, 0, 1\}$. Under this assumption, $E[W_\alpha] = E[W_\beta] = E[W_\alpha + Y]$, which implies $E[Y] = 0$. It is easily seen that the representation (72) can be rewritten as

$$\begin{aligned} Eh(W) - Eh(\pi) &= -\alpha \sum_{i=0}^{\infty} \Delta^2 f(i)\; E[Y I_{(W_\alpha - 1 \geq i+1)} + Y_+ I_{(W_\alpha - 1 = i)}] \\ &= -\alpha\; E[I_{(Y=1)} \Delta^2 f(W_\alpha - 1) + Y \Delta f(W_\alpha - 1)]. \end{aligned} \tag{73}$$

Noting that

$$\begin{aligned} |E[I_{(Y=1)} \Delta^2 f(W_\alpha - 1)]| &\leq 2\|\Delta f\|_\infty\, d_{TV}(\mathcal{L}(W_\alpha), \mathcal{L}(W_\alpha + 1)) \sup_W \{P(Y = 1 | W_\alpha)\}, \\ |E[Y \Delta f(W_\alpha - 1)]| &\leq \|\Delta f\|_\infty E|E[Y|W_\alpha]| \leq \|\Delta f\|_\infty \sqrt{\mathrm{Var}(E[Y|W_\alpha])}, \end{aligned}$$

we can immediately bound the right-hand side of (73) to obtain the following.

Proposition 4. *Suppose that* $\alpha = \beta$ *and* $EW_\alpha = EW_\beta$*. If* W_α *and* W_β *can be constructed on the same probability space such that*

$$W_\beta = W_\alpha + Y \quad \text{for some random variable } Y \text{ valued in } \{-1, 0, 1\}, \tag{74}$$

then,

$$\begin{aligned} |Eh(W) - Eh(\pi)| \leq \; & 2\alpha \|\Delta Sh\|_\infty \, d_{TV}(\mathcal{L}(W_\alpha), \mathcal{L}(W_\alpha + 1)) \sup_W \{P(Y = 1 | W_\alpha)\} \\ & + \alpha \|\Delta Sh\|_\infty \sqrt{Var(E[Y|W_\alpha])}. \end{aligned} \tag{75}$$

Clearly, if such a random variable Y takes values on a bounded set other than $\{-1, 0, 1\}$, a representation analogous to (73) may still be found, and a result analogous to Proposition 4 is available. We now apply our Proposition 4 to approximate a sum of independent indicator random variables.

Example 8. Suppose that $W = X_1 + \cdots + X_n$ is the sum of independent Bernoulli random variables with success probabilities p_i, $1 \leq i \leq n$. Brown and Xia (2001, Section 3) showed that in this case, one can improve on Poisson or binomial approximation for W by using a so–called polynomial birth–death distribution, with the choices $\alpha_j = \alpha$ and $\beta_j = \gamma j + j(j-1)$ for some constants α and γ.

We will follow that approach and choose here α and γ such that $\alpha = \beta$ and $E[\alpha_W (W + 1)] = E[\beta_W W]$. Straightforward computations then give us expressions for these parameters:

$$\gamma = \lambda^2 \lambda_2^{-1} - 1 - 2\lambda + 2\lambda_3 \lambda_2^{-1}, \quad \text{and} \quad \alpha = \gamma\lambda + \lambda^2 - \lambda_2, \tag{76}$$

where $\lambda_k = \sum_{i=1}^n p_i^k$ and $\lambda = \lambda_1 = E[W]$ (as in Section 5). Note that the parameter choices (76) are the same as those employed by Brown and Xia (2001), who based their selection on minimising the error bound obtained in their result.

To begin with, let us prove that the condition (74) is satisfied. Since the birth rate is constant (as in the Poisson case), we again have that $W_\alpha = W + 1$. Let us turn our attention to W_β. We let $W_i = W - X_i$, and $W_{i,j} = W - X_i - X_j$, $0 \leq i, j \leq n$ and observe that $W(W-1) = \sum_{1 \leq i \neq j \leq n} X_i X_j$. By the definition of W_β, we get that

$$\begin{aligned} P(W_\beta = k) &= \alpha^{-1} E\left\{[\gamma W + W(W-1)] I_{(W=k)}\right\} \\ &= \alpha^{-1}[\gamma \sum_{i=1}^n p_i P(W_i + 1 = k) + \sum_{1 \leq i \neq j \leq n} p_i p_j P(W_{i,j} + 2 = k)], \end{aligned}$$

for $1 \leq k \leq n$. In the spirit of the size-biasing construction of Section 4, we define now two random indices $T, U \in \{1, \dots, n\}$ chosen according to the distribution

$$P(T = i, U = j) = \frac{p_i p_j}{\lambda^2 - \lambda_2}, \quad i \neq j, \qquad \text{and} \qquad P(T = U = i) = 0.$$

Recall also the definition (27) of the random index V. Combining these definitions with the above, we may write

$$P(W_\beta = k) \ = \ \alpha^{-1}\gamma\lambda P(W + 1 - X_V = k) + \alpha^{-1}(\lambda^2 - \lambda_2)P(W + 2 - X_T - X_U = k),$$

for $1 \leq k \leq n$. Let $q = \alpha^{-1}\gamma\lambda$; note from (76) that $0 \leq q \leq 1$ whenever $\gamma \geq 0$. In the sequel we will assume that this is indeed the case. Introduce a Bernoulli random variable v_q with success probability q, independent of all other entries. We may then write

$$W_\beta = v_q(W + 1 - X_V) + (1 - v_q)(W + 2 - X_T - X_U) = W + 1 + Y = W_\alpha + Y,$$

where

$$Y = (1 - v_q)(1 - X_T - X_U) - v_q X_V, \tag{77}$$

Y being valued in $\{-1, 0, 1\}$ with $E[Y] = 0$, as desired.

Now, let us evaluate the bound (75). First, we need a bound on the solution f of the Stein equation in this situation. By Theorem 2.10 of Brown and Xia (2001), one knows that

$$\sup\{\|\Delta Sh\|_\infty : h(j) = I_{(j \in B)},\ B \subseteq \mathbb{Z}^+\} \leq \alpha^{-1}. \tag{78}$$

Further, W being a sum of independent indicators, one has (from Barbour and Jensen (1989, Lemma 1))

$$d_{TV}(\mathcal{L}(W), \mathcal{L}(W+1)) \ \leq \ \frac{1}{2\sqrt{\sum_{i=1}^n p_i(1-p_i)}}. \tag{79}$$

Finally, consider the two conditional terms in (75). Note from (77) that $Y = 1$ if and only if $v_q = X_T = X_U = 0$, so that

$$\begin{aligned} P(Y = 1|W) \ &= \ (1 - q)P(X_T = X_U = 0|W) \ = \ (1 - q)E[(1 - X_T)(1 - X_U)|W] \\ &= \ \alpha^{-1} \sum_{1 \leq i \neq j \leq n} p_i p_j E[(1 - X_i)(1 - X_j)|W]. \end{aligned}$$

This probability takes its greatest value when $W = 0$, with $E[(1 - X_i)(1 - X_j)|W = 0] = 1$ for all i and j. Hence,

$$\sup_W\{P(Y = 1|W)\} = \alpha^{-1} \sum_{1 \le i \ne j \le n} p_i p_j = \alpha^{-1}(\lambda^2 - \lambda_2). \tag{80}$$

Now, let $\|Z\| = (E[Z^2])^{1/2}$ be the L_2 norm for any random variable Z. Since $T =_d U$ and $E[Y] = 0$, we write

$$E[Y|W] = -q(E[X_V|W] - E[X_V]) - 2(1 - q)(E[X_T|W] - E[X_T]),$$

and thus

$$\begin{aligned}
\sqrt{\operatorname{Var}(E[Y|W])} &= \|E[Y|W]\| \\
&\le q \sum_{j=1}^{n} \|E[X_j|W] - E[X_j]\| P(V = j) \\
&\qquad + 2(1 - q) \sum_{j=1}^{n} \|E[X_j|W] - E[X_j]\| P(T = j) \\
&\le (q + 2(1 - q)) \max_{1 \le j \le n} \sqrt{\operatorname{Var}(E[X_j|W])}.
\end{aligned}$$

When $p_j = p$ for $j = 1, \ldots, n$, $E[X_j|W] = W/n$ and so the bound becomes the equality

$$\sqrt{\operatorname{Var}(E[Y|W])} = (2 - q)\sqrt{\operatorname{Var}(W/n)}. \tag{81}$$

Inserting (78), (79), (80) and (81) in (75) then provides the following bound:

$$d_{TV}(\mathcal{L}(W), \mathcal{L}(\pi)) \le \frac{p}{(1 - p)\sigma} + \frac{(2 - q)\sigma}{n} = O(p/\sqrt{\lambda}),$$

where $\sigma^2 = \operatorname{Var}(W) = np(1 - p)$.

By exploring the explicit structure of the auxiliary variable Y, it is possible to derive better bounds. Throughout this part we let $\bar{a} = 1 - a$ for any $a \in \mathbb{R}$ and $\sigma_k = \sqrt{\sum_{i=k+1}^{n} \rho_i}$, where ρ_i is the ith largest number of $p_1(1 - p_1), \ldots, p_n(1 - p_n)$. From Barbour and Jensen (1989, Lemma 1) we have that for all $i, j = 1, \ldots, n$ and $i \ne j$,

$$2d_{TV}(\mathcal{L}(W_i), \mathcal{L}(W_i + 1)) \le \sigma_1^{-1} \quad \text{and} \quad 2d_{TV}(\mathcal{L}(W_{i,j}), \mathcal{L}(W_{i,j} + 1)) \le \sigma_2^{-1}.$$

Notice that, from representation (77),

$$I_{(Y=1)} = \bar{v}_q \bar{X}_T \bar{X}_U \ , \ I_{(Y=-1)} = v_q X_V + \bar{v}_q X_T X_U. \tag{82}$$

The derivations below are based on the conditional independence of X_T and W_T, given T and similarly X_U and W_U, given U and X_V and W_V, given V. By substituting (82) in (73), integrating with respect to v_q, separating linear and quadratic terms and noticing that $T =_d U$, we derive, after some simple calculations,

$$\begin{aligned}
I &= Eh(W) - Eh(\pi) \\
&= -\alpha E[\bar{v}_q \bar{X}_T \bar{X}_U \Delta f(W+1)] + \alpha E[(v_q X_V + \bar{v}_q X_T X_U)\Delta f(W)] \\
&= -\alpha \bar{q} E[X_T X_U \Delta^2 f(W)] \\
&\qquad + 2\alpha \bar{q} E[X_T \Delta^2 f(W)] \\
&\qquad - \alpha(\bar{q} E[\Delta f(W+1)] - E[(2\bar{q} X_T + q X_V)\Delta f(W)]) \\
&= I_1 + I_2 + I_3.
\end{aligned}$$

Using the conditional independence of $W_{T,U}$ and X_T, X_U given T and U, the first term I_1 is bounded by

$$\begin{aligned}
|I_1| &= \alpha \bar{q} \big| E E[X_T X_U | T, U] E[\Delta^2 f(W_{T,U} + 2)] \big| \\
&\leq 2\alpha \bar{q} \|\Delta f\|_\infty E[X_T X_U] \max_{i \neq j} \big\{ d_{TV}(\mathcal{L}(W_{i,j}), \mathcal{L}(W_{i,j}+1)) \big\} \;\leq\; \frac{\lambda_2^2 - \lambda_4}{\alpha \sigma_2}.
\end{aligned}$$

By conditioning on T,

$$\begin{aligned}
|I_2| &= 2\alpha \bar{q} \big| E E[X_T | T] E[\Delta^2 f(W_T + 1)] \big| \\
&\leq 4\alpha \bar{q} \|\Delta f\|_\infty E[X_T] \max_i \big\{ d_{TV}(\mathcal{L}(W_i), \mathcal{L}(W_i + 1)) \big\} \;\leq\; \frac{2(\lambda \lambda_2 - \lambda_3)}{\alpha \sigma_1}.
\end{aligned}$$

To bound I_3, we first notice that since $E[Y] = 0$,

$$\bar{q} = 2\bar{q} E[X_T] + q E[X_V].$$

Thus,

$$\begin{aligned}
|I_3| &= \big| 2\alpha \bar{q} E\big\{ X_T \big(E[\Delta f(W_T + 1) | T] - E[\Delta f(W_T + X_T + 1)] \big) \big\} \\
&\qquad + \alpha q E\big\{ X_V \big(E[\Delta f(W_V + 1) | V] - E[\Delta f(W_V + X_V + 1)] \big) \big\} \big| \\
&\leq 2\alpha \big\{ 2\bar{q} E[\{E(X_T | T)\}^2] \\
&\qquad\qquad + q E[\{E(X_V | V)\}^2] \big\} \|\Delta f\|_\infty \max_i \big\{ d_{TV}(\mathcal{L}(W_i), \mathcal{L}(W_i + 1)) \big\} \\
&\leq \frac{2(\lambda \lambda_3 - \lambda_4)}{\alpha \sigma_1} + \frac{\gamma \lambda_3}{\alpha \sigma_1}.
\end{aligned}$$

By combining the bounds on I_1, I_2 and I_3 we derive the following.

Proposition 5. *With W and π as above,*

$$d_{TV}(\mathcal{L}(W),\mathcal{L}(\pi)) \;\leq\; \frac{\lambda_2^2-\lambda_4}{\alpha\sigma_2}+\frac{2(\lambda\lambda_2-\lambda_3)}{\alpha\sigma_1}+\frac{2(\lambda\lambda_3-\lambda_4)}{\alpha\sigma_1}+\frac{\gamma\lambda_3}{\alpha\sigma_1}. \tag{83}$$

Let us conclude by comparing our result with that of Brown and Xia (2001, Theorem 3.1), who obtain

$$d_{TV}(\mathcal{L}(W),\mathcal{L}(\pi)) \;\leq\; \frac{\gamma\lambda_3}{\alpha\sigma_1}+\frac{2\lambda\lambda_2}{\alpha\sigma_2}. \tag{84}$$

When $p_i = p \to 0$ for each i and $\lambda \to \infty$, both the bounds (83) and (84) are asymptotically equivalent to $3p^2/\sqrt{\lambda}$.

Acknowledgements

Fraser Daly gratefully acknowledges the financial support of the Schweizerische Nationalfonds, and thanks the Belgian Fonds National de la Recherche Scientifique for support during a visit to the Université Libre de Bruxelles. Thanks are also due to Andrew Barbour for several useful discussions.

References

Arratia, R., Goldstein, L. and Gordon, L. (1989). Two moments suffice for Poisson approximations: the Chen–Stein method. *Ann. Prob.* **17**, 9–25.

Barbour, A. D. and Chen, L. H. Y., editors (2005). *An Introduction to Stein's Method*. Lect. Notes Ser., Inst. Math. Sci. 4, Natl. Univ. Singap., Singapore.

Barbour, A. D., Holst, L. and Janson, S. (1992). *Poisson Approximation*. Oxford Univ. Press, Oxford.

Barbour, A. D. and Jensen, J. L. (1989). Local and tail approximations near the Poisson limit. *Scand. J. Statist.* **16**, 75–87.

Barbour, A. D. and Pugliese, A. (2000). On the variance-to-mean ratio in models of parasite distributions. *Adv. Appl. Prob.* **32**, 701–719.

Barbour, A. D. and Xia, A. (1999). Poisson perturbations. *ESAIM Prob. Stat.* **3**, 131–150.

Brown, T. C. and Phillips, M. J. (1999). Negative binomial approximation with Stein's method. *Methodol. Comput. Appl. Prob.* **1**, 407–421.

Brown, T. C. and Xia, A. (2001). Stein's method and birth-death processes. *Ann. Prob.* **29**, 1373–1403.

Čekanavičius, V. and Vaĭtkus, P. (2001). Centred Poisson approximation via Stein's method. *Lith. Math. J.* **41**, 319–329.

Chatterjee, S., Diaconis, P. and Meckes, E. (2005). Exchangeable pairs and Poisson approximation. *Probab. Surv.* **2**, 64–106.

Chen, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Prob.* **3**, 534–545.

Ehm, W. (1991). Binomial approximation to the Poisson binomial distribution. *Stat. Prob. Lett.* **11**, 7–16.

Erhardsson, T. (2005). Stein's method for Poisson and compound Poisson approximation. In Barbour, A. D. and Chen, L. H. Y., editors (2005). *An Introduction to Stein's Method*. Lect. Notes Ser., Inst. Math. Sci. 4, Natl. Univ. Singap., Singapore, 61–113.

Goldstein, L. and Rinott, Y. (1996). Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Prob.* **33**, 1–17.

Holmes, S. (2004). Stein's method for birth and death chains. In Diaconis, P. and Holmes, S., editors (2004). *Stein's Method: Expository Lectures and Applications*. IMS Lect. Notes 46, Beachwood, Ohio, 45–67.

Johnson, N. L., Kotz, S. and Kemp, A. W. (1992). *Univariate Discrete Distributions*. Wiley, New York.

Kamae, T., Krengel, U. and O'Brien, G. L. (1977). Stochastic inequalities on partially ordered spaces. *Ann. Prob.* **5**, 899–912.

Lefèvre, C. and Utev, S. (1996). Comparing sums of exchangeable Bernoulli random variables. *J. Appl. Prob.* **33**, 285–310.

Papadatos, N. and Papathanasiou, V. (2002). Approximation for a sum of dependent indicators: an alternative approach. *Adv. Appl. Prob.* **34**, 609–625.

Peköz, E. (1996). Stein's method for geometric approximation. *J. Appl. Prob.* **33**, 707–713.

Phillips, M. J. and Weinberg, G. (2000). Non-uniform bounds for geometric approximation. *Stat. Prob. Lett.* **49**, 305–311.

Reinert, G. (2005). Three general approaches to Stein's method. In Barbour, A. D. and Chen, L. H. Y., editors (2005). *An Introduction to Stein's Method*. Lect. Notes Ser., Inst. Math. Sci. 4, Natl. Univ. Singap., Singapore, 183–221.

Röllin, A. (2005). Approximation of sums of conditionally independent variables by the translated Poisson distribution. *Bernoulli* **11**, 1115–1128.

Röllin, A. (2007). Translated Poisson approximation using exchangeable pair couplings. *Ann. Appl. Prob.* **17**, 1596–1614.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders.* Springer, New York.

Stein, C. (1986). *Approximate Computation of Expectations.* IMS Lect. Notes, Monogr. Ser. 7, Hayward, California.

Wang, X. and Xia, A. (2008). On negative binomial approximation to k-runs. *J. Appl. Prob.* **45**, 456–471.